DOCUMENT RESUME

ED 288 922 UD 025 895

AUTHOR	Slavin, Robert E.; Madden, Nancy A.
TITLE	Effective Classroom Programs for Students at Risk.
INSTITUTION	Center for Research on Elementary and Middle Schools, Baltimore, MD.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Apr 87
GRANT	OERI-G-86-0006
NOTE	49p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE	Speeches/Conference Papers (150) -- Information Analyses (070) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Compensatory Education; *Continuous Progress Plan; Elementary Education; Grouping (Instructional Purposes); *High Risk Students; *Individualized Instruction; Program Effectiveness; Remedial Instruction
IDENTIFIERS	*Cooperative Learning

ABSTRACT

This paper examines research on classroom programs for elementary school students who are at risk for learning problems. The full range of alternative classroom organization models designed to meet the needs of low-achieving or heterogeneous classes is explored in an attempt to discover which type of program is most effective and why. The goal of this study is to determine how the education needs of all students can be met by fundamentally restructuring the regular classroom, as opposed to adding on services outside of the regular classroom. Therefore this review focuses on the specific, comprehensive programs beneficial to the achievement of students who are at risk and that are replicable at other schools. Research demonstrates that effective classroom programs accommodate instruction to individual needs while maximizing direct instruction, and assess student progress frequently through a structured hierarchy of skills. Two categories of programs emerged as particularly effective: continuous progress and cooperative learning. An extensive list of references is included. Descriptions of several types of successful programs are appended. (PS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

11-19-87 4U

ED288922

Effective Classroom Programs For Students at Risk

Robert E. Slavin

Nancy A. Madden

Center for Research on Elementary and Middle Schools

Johns Hopkins University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the Annual Convention of the American Educational Research Association, Washington, D.C., April, 1987. This paper was written under funding from the Office of Educational Research and Improvement (Grant No. OERI-G-86-0006). However, the opinions expressed are those of the authors and do not necessarily reflect OERI positions or policy.

UD025895

BEST COPY AVAILABLE

In recent years, there has developed a growing dissatisfaction with the educational services provided to students who are "at risk" for developing learning problems or school failure. Much of this dissatisfaction has centered on the practice of pulling low achieving students out of class for remedial instruction, usually provided as part of the Federal Chapter I (formerly Title I) program. Some researchers have found that the more time students spent in pull-out programs, the <u>less</u> they learned (Coulson et al., 1977; Glass & Smith, 1977). Chapter I pull-outs have been criticized on the basis that instruction in the pull-out program is rarely well integrated with that provided by the regular classroom teacher (Kaestle & Smith, 1982; Johnston, Allington, & Afflerbach, 1985). Also, much time is lost in transitions between regular and pull-out settings (Allington, in press), and pull-outs rarely increase the total instruction provided to students (Vanecko, Ames, & Archambault, 1980; Kimbrough & Hill, 1981). Similar problems of coordination and of supplanting of regular classroom instruction have been noted for special education pull-outs (Sargent, 1981) and when students qualify for special education <u>and</u> remedial services, these problems multiply (Birman, 1981).

While the problems of pull-outs are well recognized, <u>alternatives</u> to pull-outs have their own drawbacks. Having the remedial teacher work in a corner of the regular classroom, as is typical in in-class Chapter I models, provides no guarantee that coordination with the regular program will be enhanced (see Lee & Rowan, 1986). "In-class" models are often structurally identical to pull-out programs except that the remedial services are provided in a different location. Research comparing pull-out and in-class Chapter I/Title I models has found few differences in achievement effects (Archambault, 1987).

One point on which there is agreement among virtually all reviewers of research on Chapter I and special education programs for students at risk is that program quality is far more important than the setting in which the program is implemented (Leinhardt & Pallay, 1983; Madden & Slavin, 1983; Archambault, 1987). Clearly, there can be effective pull-out programs and effective in-class programs to meet the needs of at-risk students. This paper examines research on classroom programs for elementary school students who are at risk for learning problems. The full range of alternative classroom organization models designed to meet the needs of low-achieving or heterogeneous classes is explored; the principal question this paper asks is how can the educational needs of all students be met by fundamentally restructuring the regular classroom, as opposed to adding on services outside of the regular classroom. A companion paper (Madden & Slavin, 1987) reviews research on effective pull-out programs for elementary students, and a third paper in this series by Nancy Karweit (1987) discusses effective programs for preschool and kindergarten students.

Defining "At Risk"

The definition of "at risk" applied in this paper is deliberately vague. Many factors may identify a student as being at risk for school failure; among these are low socioeconomic status, low measured ability, learning disabilities, or perhaps most importantly, learning problems early in the schooling experience. While each of these risk factors is conceptually distinct, all but the first (low SES) are difficult to distinguish in practice (see Ysseldike & Algozzine, 1982). With young students, low socioeconomic status may be used as a very rough means of identifying groups

ERIC

of students who are likely, on average, to experience learning problems, but of course many low-SES students are quite successful in school. However, by the end of the first or second grade it is often clear that some students, regardless of SES, are beginning a process of falling further and further behind. With the exception of a few easily remediable conditions (such as poor eyesight or hearing), the fact that students are not achieving adequately in the early grades, for whatever reason, is the most important diagnostic indication that a students is at risk. Students who had early school problems are heavily overrepresented among school failures (Lloyd, 1978), dropouts (Stroup & Robins, 1972), delinquents (Kelly, Veldman, & McGuire, 1964), and others who ultimately experience serious problems as they proceed through the school years.

Scope of the Review

The focus of this review is on specific, comprehensive programs that may be beneficial for the achivement of students who are at risk for learning problems. A program is defined here as a set of procedures intended to be implemented as a total package and capable of being replicated by others. These features distinguish a program from a set of variables. For example, there is evidence that such variables as strong principal leadership, high expectations, and high time on-task are related to student achievement gains (see, for example, Brophy & Good, 1986). However, these variables do not in themselves constitute a program; a program would lay out a specific set of procedures that might incorporate these variables (and others) but would do so in a structured and replicable format. The word "replicable" is particularly important in the definition of a program. It is often the case

ERIC

that a particular school or district has considerable success with at-risk students, but the success is due to the unusual skill or devotion of a set of individual principals or teachers rather than to any particular set of practices that can be easily transmitted to others.

Review Procedures

This review used a set of procedures adapted from best-evidence synthesis (Slavin, 1986), a review method which combines features of meta-analytic and traditional narrative reviews. In essence, a priori inclusion criteria relating to internal and external validity were established and a broad literature search was conducted to identify studies which met those criteria. When possible, achievement effects of the qualifying studies were characterized in terms of effect size, the difference between experimental and control means divided by the control groups' standard deviation (Glass, McGaw, & Smith, 1981). However, the strengths and weaknesses of the individual evaluations are described to set the effects in context. These procedures are described in more detail in the following sections.

Literature Search. Every effort was made to search the literature on instructional programs for elementary-aged students (Grades 1-6) as broadly as possible. Academic journals yielded a few relevant studies, but research on comprehensive instructional programs is rarely reported in these journals. The most fruitful source of information by far was reports submitted to the Joint Dissemination Review Panel (JDRP), a U.S. Department of Education panel which reviews evaluations of programs originally supported by federal funds. Programs whose effects are certified as valid by the JDRP are eligible for funding and dissemination through the National

Diffusion Network (NDN). What makes the JDRP submissions such a rich source of information for identifying effective programs is that with few exceptions the programs submitted to the JDRP were designed to be replicable; the main reason for a school district or research and development organization to go through JDRP review is to obtain funds to nationally disseminate a program, so data on programs that could not be easily transported to a new location would be unlikely to be submitted to the JDRP. Also, the JDRP review process demands enough data to enable reviewers to determine program effects. The submissions for more than 300 programs which successfully passed JDRP review were obtained from the Department of Eduacation. Also, the Northwest Regional Education Laboratory recently published an _Effective Compensatory Education Sourcebook_ (Griswold, Cotton, & Hansen, 1986) which listed 116 district Chapter I programs nominated and selected as being particularly effective. Data and program descriptions were requested from these districts. Finally, ERIC documents, journal articles, and other reports were obtained in a general literature search relying in particular on references from other reviews (e.g., Ellson, 1987).

Substantive Inclusion Criteria In order to be considered germane to the scope of this review, studies had to conform to a broad set of criteria:

1. The programs evaluated had to be directed toward increasing the reading and/or mathematics achievement of students in grades 1-6. Programs which were designed primarily for use in kindergarten or preschool are discussed by Karweit (1987), while those which overlapped the elementary grades (e.g., K-2, K-6, K-12) are included here.

2. The programs had to be implemented in regular classrooms. This excludes self-contained special education programs. Pull-out and tutoring programs are reviewed by Madden & Slavin (1987).

3. The programs had to be applicable to the education of at-risk students. This criterion was interpreted very broadly. Even if programs did not present evidence that they had been successfully evaluated in schools containing disproportionate numbers of at-risk students, they were included if they could in principal be applied to such students. However, a lack of data on effects for at-risk students is noted as a drawback in descriptions of effective programs.

4. Programs specifically designed for non-English speakers are not included in this review.

Methodological Inclusion Criteria. Unfortunately, the methodological quality of most evaluations of instructional programs is very low. Many evaluations compare the achievement of students in experimental programs to that of students in "comparable" schools, with no evidence given that the comparison schools were in fact comparable before the treatments were administered. Others lack even this much control, claiming that since students achieved above "grade level expectations," the program must be successful.

Perhaps the most common reason for invalidity of evaluation design is the use of fall-to-spring gains in percentiles or, more commonly, normal curve equivalents (NCE's). A normal curve equivalents is a standard score with a mean of 50, a standard deviation of about 21, and a range of 1 to 99. NCE's

are therefore similar to percentile scores, but they are preferable because they use an equal-interval scale. In recent years, districts receiving Chapter I funds have been encouraged to report fall-to-spring gains in NCE's for their Chapter I recipients, which helps explain the popularity of this design.

The rationale behind the fall-to-spring design is sensible in theory. If students score at, say, a mean NCE of 40 in the fall and 50 in the spring, then they have apparently gained relative to the national norming group, which should have remained at the same level on average. However, in recent years it has become apparent that this assumption is incorrect. Among Chapter I populations, NCE gains of about 8 points are routinely made from fall to spring (see Gabriel et al., 1985; Keesling, 1984). However, by the following fall, these differences have generally disappeared. The fall-to-spring gain is probably a statistical artifact, not a true effect of Chapter I. For example, Chapter I-eligible students who do not actually receive Chapter I services make similar gains (Carter, 1984; Granick et al., 1984). Almost all of the "exemplary" Chapter I programs cited by Griswold et al. (1986) which provided multi-year data showed a pattern of fall-to-spring gains followed by equal spring-to-fall declines; the fall-to-fall or spring-to-spring gains were essentially nil over as many as five years. National data reported by Gabriel et al. (1985) indicate that the fall-to-spring effect is largest in the younger grades, diminishing from almost 9 NCE's in grade 2 to 4.5 in grade 12.

Unfortunately, the cause of the fall-to-spring artifact is unknown (see Kennedy and Birman, 1986), and it is unlikely that the gains are equal

across tests, types of students, or other factors. Therefore, it is not even possible to assume that because a particular district reports NCE gains much more than eight that this is evidence of an effective program. There are examples of school districts in which fall-to-spring gains of 16 NCE's or more are reported each year, yet no growth from fall-to-fall or spring-to-spring is evident.

Program evaluations which use fall-to-fall or spring-to-spring NCE or percentile gains as a criterion for program effects are better than fall-to-spring studies, because the expected gains for annual testing are much smaller, in the range of 1-3 NCE's (Gabriel et al, 1983). Yet these have serious drawbacks as well. Some evaluations report annual gains for successive cohorts of students at a particular grade level (e.g., third graders in 1978-80 vs. third graders in 1980-81). Cohort differences can be affected by many extraneous factors, such as changes in district testing or promotion procedures. However, this design is preferable to a comparison of this year's gain to last year's for all students in a grade level, which confounds programs effects with retention/promotion, selection into or out of Chapter I or other special programs, and other factors. Comparing last year's gains to this year's for promoted students only is a much better procedure, but is rarely used; including non-promoted students inflates estimates of annual gain because these students are being compared ag in to the same grade-level norms after a year of instruction and growth.

The inclusion criteria applied in the present review follow from the above discussion. The purpose of these criteria is to extract from a large, diverse, and messy literature the most convincing evidence available at

ERIC

present on effective classroom programs for students at risk. The criteria are as follows.

1. Convincing evidence of effectiveness had to be presented. Unfortunately, the nature of existing program evaluation data and of the sources of reports on such data precludes identification of programs which are not effective. The JDRP and examplary Chapter I programs only come to our attention because they were felt to be outstanding. A pattern of fall-to-spring gains cannot be used as evidence of effectiveness, but neither is it convincing evidence of ineffectiveness. In a usual review of the literature on experimental programs, effective and ineffective programs would be compared in an attempt to identify consistent characteristics of the more effective ones (see Cooper, 1984; Light & Pillemer, 1984; Slavin, 1986). However, in the present case, the review focuses on those programs with convincing evidence of effectiveness, as distinct from others which may or may not be equally effective but do not present convincing data. A list of programs considered for inclusion but not included and reasons for exclusion is presented later in this paper.

The remaining methodological inclusion criteria define what is meant by "convincing evidence of effectiveness."

2. Eval_ ons had to use control group designs with random assignment to groups and/or convincing evidence that comparison groups were initially equivalent in achievement. Studies which showed spring-to-spring gains of at least seven NCE's were also included, but in a separate category. This is about one-third of a population standard deviation (approximately 21), but more like half of the standard deviation typical of Chapter I students,

who have a more restricted range of scores. One important category of programs excluded from the set of effective models is those which were involved in the Abt Follow-Through evaluation and were found to be no more effective than control groups overall, even though they were found to be effective in one or more sites. In many cases, data from the successful sites passed JDRP review, as the JDRP procedures do not require that data from all sites be submitted or considered.

3. Evaluations had to use standardized, broadly based measures of reading and/or mathematics achievement. This excluded a small number of studies which used experimenter-made measures or only assessed a narrow set of reading or mathematics skills.

4. The duration of evaluations had to be at least one semester (16 weeks). Almost all JDRP and Chapter I evaluations cover at least a one-year period, but many evaluations of promising programs which have appeared in academic journals have involved much shorter treatment durations. To be considered a practical evaluation of a replicable program, a duration of one semester is a minimal requirement.

Categories of Effective Models

The principal results of the search for instructional methods with convincing evidence of positive effects are summarized in Tables 1-3. In each table, programs evaluated in studies which employed matched or (rarely) randomized control groups are listed first, with associated effect sizes. These effect sizes should be interpreted very cautiously and should not be compared across methods, but are included to characterize the relative

magnitudes of effects on different outcomes or for different subpopulations within programs.

The effective models are discussed according to categories of programs. Surprisingly, among the dozens of programs evaluated for possible inclusion, programs which presented convincing data fell into only three categories: continuous progress, individualized instruction, and cooperative learning. There are defined below.

Continuous Progress

By far the largest number of programs with convincing evidence of effectiveness are those falling in the category "continuous progress." The individual programs in this category vary in many particulars, but features common to all include the following.

1. Students procede through a well-specified hierarchy of skills. They are tested at each level to determine their readiness to move on to the next skill. Special procedures are established to assist students who fail to pass regular mastery assessments. These may include corrective instruction in small groups, tutoring, assignment to different groups, or special materials or activities. In these respects continuous progress programs resemble mastery learning and they are often described as such. However, continuous progress programs are fundamentally different from the group-based mastery learning programs more commonly seen in elementary and secondary schools (see Slavin, in press a). Group-based mastery learning is discussed later in this paper.

2. Careful records are kept of each student's progress through the curriculum. These data are used to make grouping, remediation, and other decisions.

3. Most instruction is delivered by teachers to groups of students at the same instrucitonal level. This is primarily what differentiates continuous progress from individualized models. In continuous progress models, students progress at their own rates, but are primarily instructed by teachers rather than by programmed or other individualized materials. In these methods, students are constantly grouped and regrouped according to their levels of reading and/or math skills; this grouping is often done across grade lines, as in nongraded programs (see Goodlad & Anderson, 1963).

Individualized Instruction

1. Students work primarily on programmed or other individualized materials, with teachers working mostly with individuals rather than groups of students. This is the primary difference between continuous progress and individualized models; many continuous progress models describe themselves as "individualized," but do not rely on self-instructional materials.

2. As in continuous progress programs, individualized models keep careful records of student progress through a structured, hierarchical set of learning objectives.

Cooperative Learning

Although there are many methods based on principles of cooperative learning (See Sharan, 1980; Slavin, 1983), only two have been evaluated over

periods of at least one semester in terms of effects on standardized tests of reading and math: Team Assisted Individualization (TAI; Slavin, 1985) and Cooperative Integrated Reading and Composition (CIRC; Stevens, Madden, Slavin, & Farnish, in press). In addition to structuring cooperation among students, both methods also incorporate elements of continuous progress as well. The elements in common to TAI and CIRC are as follows.

1. Students work in 4-5 member mixed-ability learning teams and receive certificates or other recognition based on the learning of all group members. Students both help one another learn and assess one another's skills in preparation for tests or teacher assessments which will be taken without teammate help.

2. Teachers instruct students (drawn from the different teams) who are at the same level in a hierarchy of skills. Most information comes from the teachers, not from peers or materials; teammates help one another master skills, but do not provide instruction to one another.

3. Students are frequently assessed, first by teammates and then by the teacher. Specific corrective procedures are provided for students who do not meet a preset level of mastery.

Table 1 Here

ERIC

Characteristics and Outcomes of Effective Models

Continuous Progress

A total of eleven programs were categorized as continuous progress models with convincing evidence of effectiveness. Seven of these used control group designs.

Distar. Distar (Becker & Carnine, 1980) is a program developed at the University of Oregon. Distar is unusual, even within the range of continuous progress programs. It provides teachers with very specific scripts to use in teaching reading and math and trains teachers in very specific methods, down to the level of how to use hand signals to elicit student responses and how to call on students. In describing the program, Distar's developers tend to emphasize the sequential, hierarchical curriculum design, direct instruction and rapid pace, and high frequency of student reponses (Becker & Carnine, 1980). However, what defines Distar as a continuous progress model is that students are taught in small groups that are homogeneous in skill level, are assessed frequently on their progress through a well-defined skill hierarchy, and are regrouped if necessary according to the results of these assessments.

The evidence supporting the effectiveness of Distar for increasing student performance in certain skill areas is very strong. Distar was the only one of the nine major programs evaluated in the national Planned Variation Follow Through studies conducted by Abt Associates (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977) to have consistently positive effects on the achievement of disadvantaged students. The effect size

estimates which appear in Table 1 are from the Abt evaluation, pooling across cohorts (see Kennedy, 1978). As the table indicates, however, the effects of Distar appeared primarily on language and math computations tests, not on such higher-order skills as reading comprehension or math problem solving. Even on these scales, though, Distar students scored higher than all other Follow Through models. Note that these effects were brought about over three or four years of implementation, as the Abt evaluation focused on students who were in their respective programs from kindergarten or first to third grade.

What makes the Abt evaluation results particularly impressive is that this evaluation involved many sites, matched control groups, and consistent measurement across non-Follow Through and various Follow Through treatments. In reviewing instructional programs for this paper, there was always a nagging question of whether reports of successful evaluations sent to the JDRP or published in a journal are anomalies, perhaps one or two sites out of dozens in which the program was evaluated. In fact, in several cases sites which had positive results successfully passed JDRP review even though they were using methods which were unsuccessful at most other sites. The effects for Distar summarized in Table 1 are averaged across all sites which used the program, successfully or not.

Recent research on Distar has found that the program's effects can be long-lasting. Following up students from an inner-city New York elementary school, Distar students exceeded control in high school graduation; about 55% of the former Distar students graduated, compared to 34% of control students (Meyer, 1984).

Despite the well-documented and widely acknowledged positive effects of Distar, there has also been much criticism of this approach, principally on the basis of its use of scripted lessons and a perceived focus on rote skills rather than higher order, learning-to-learn skills (see, for example, Calfee, 1986). Many teachers and administrators resist Distar for similar reasons. Given this reality, it is important to note that Distar is only one of several successful continuous progress models, is the only one to use scripted lessons, and is one of only two (with ECRI) to emphasize the highly organized, teacher-centered classroom organization that many teachers find offensive.

U-SAIL. U-Sail (for Utah System Approach to Individualized Learning) was developed and evaluated in Utah. It is a continous progress program quite different from Distar. In it, students procede through a hierarchical sequence of objectives at their own rates. Most instruction is given to small groups, but some individual work is included. While the teacher is instructing skill-level groups, other students are given independent, exploratory activities, with a particular emphasis on independent reading (See Hales, 1983 a, b).

The JDRP submission for U-SAIL presents data from only one elementary school in suburban Davis County, Utah. Evidence is provided to show that the school was similar to the control school for three years before U-SAIL was begun, but gained more than the control school afterwards. The data in Table 1 represent the degree to which U-SAIL students out-performed control students in reading and math from grades 3-5. Unfortunately, we have no idea how successful the program has been in other schools, or whether the

active involvement of the developer is needed to make the program successful. Also, no evidence is given that U-SAIL can be effective with disadvantaged or low-achieving students.

PEGASUS. This program was apparently the first to pass JDRP review, as it was given the first JDRP number. PEGASUS is a classic continuous progress program. It organizes the reading program into 17 levels, spanning grades K-8. Students progress through these levels at their own rates, but are taught in groups appropriate to their current levels. Within each of the levels is a continuum of skills which students must master.

PEGASUS was originally developed in Tuscaloosa, Alabama, but data could only be obtained for evaluation of a "turnkey" site in rural Illinois. In this evaluation, schools in Bureau County were matched with similar schools in La Salle County for a one-year study. Results indicated significantly greater growth in reading for PEGASUS students at most grade levels on vocabulary as well as reading comprehension measures. However, while some of the students were poor, results for these students were not separately reported, so generalization to disadvantaged groups may not be warranted.

ECRI. The Exemplary Center for Reading Instruction, or ECRI was, like U-SAIL, originally developed in Utah. However, it is quite unlike U-SAIL in procedures, more resembling Distar in classroom organization, rapid instructional pace, detailed and specific instructions for teachers, and frequent assessment of student progress. In ECRI, students are usually assigned to three reading groups. They receive instruction in their groups, and then work on materials at their own rates. ECRI emphasizes teaching specific word attack skills and proceding through large numbers of words in a short time.

Most evaluations of ECRI presented as part of the JDRP submission used fall-to-spring designs. One study, in a relatively low SES school in a suburb of Cincinnati, used a control group design, and the results of this study were used to compute the effect sizes shown in Table 1. Also, spring-to-spring data from a mostly low SES, all black school in rural Louisiana and from a Chapter I school in suburban Englewood, Ohio suggest that ECRI can be effective with disadvantaged and low-achieving students.

Project INSTRUCT. Like PEGASUS, Project INSTRUCT is a classic continuous progress program. It was developed and evaluated in Lincoln, Nebraska. In the program, students are grouped according to skill levels, and may procede through a hierarchy of skills at their rates. An evaluation of the program by Novak (1975) compared schools which successfully implemented Project INSTRUCT to matched control schools, and found small positive effects in reading and math.

GEMS. Goal-Based Educational Management System, or GEMS, is a diagnostic-prescriptive reading program developed in Utah. In this program, students procede at their own rates through 200 skill levels covering grades K-12. Each unit has a pretest and a posttest. According to placement tests or pretests, students are placed in appropriate instructional groups. Teachers use a variety of teaching strategies, following which students must attain a score of at least 80% to exit the unit. Those who do not achieve this score are given alternate materials and enough time as they need to ultimately pass. A computer management system helps teachers keep track of student progress.

An evaluation in rural Jordan, Utah found that GEMS students generally exceeded control in vocabulary, although differences in reading comprehension were small and were not found in grades 1 and 2. No evidence is given to suggest that the program might be effective with disadvantaged students.

Early Childhood Preventative Curriculum. This program is quite different from the others in Table 1 in that it is intended to be used intensively with high-risk first graders. These students are put into a special class during first grade and experience an individualized diagnostic-prescriptive program, in which students' individual strengths and weaknesses are identified and students are allowed to procede at their own rates. Most instruction is given in small, skill-level groups.

ECPC was developed and evaluated in Miami, Florida. The evaluation compared the program to matched as well as random controls. Overall, positive effects were found on the Paragraph Meaning scale of the Stanford Achievement Test and (to a lesser extent) on the WRAT.

Continuous Progress Programs Evaluated Using Year-to-Year Gains. In addition to the programs listed above, four additional continous progress models presented convincing evidence of year-to-year gains: Weslaco Individualized Reading/Language Arts Instruction and Staff Development Process (WILASD), Conceptually Oriented Mathematics Curriculum (COMP), Coordinated Learning Integration -- Middlesex Basics (CLIMB), and Outcomes-Driven Developmental Model (ODDM). The procedures used in these models generally fall within the range represented among the other continous progress programs.

A few additional continuous progress programs passed by the JDRP might also be effective, but did not present adequate evidence. For example, Intensive Reading Instructional Teams (IRIT; JDRP No. 74-11) presented only fall-to-spring percentile gains. The Cranston Comprehensive Reading Program (JDRP NO. 82-28) used a regression-discontinuity evaluation design which showed an increase in student scores the year the program was begun, but the district also changed test forms at the same time, making the gains difficult to ascribe to the program with any confidence.

Table 2 Here

Individualized Instruction

Table 2 lists individualized instruction programs which presented convincing evidence of effectiveness. Only three programs fall into this category and of these, only one used a control group design.

Matteson Four-Dimensional Reading Program. This program, developed and evaluated in a Chicago suburb, provides students with individualized learning packets through which students procede at their own rates. Some small- and large-group activities are used, and the individualized work occupies only 40-60% of the total reading period, so the Matteson model cannot be called a "pure" individualized program. An evaluation comparing matched schools found that fourth graders in the Matteson program gained more than control students in Stanford Achievement Test scores over a three-year period. However, no evidence is given that would suggest that this program could be successfully aplied to at-risk students.

AIRS. Andover's Individualized Reading System (AIRS) is another self-paced instructional model developed and evaluated in a well-to-do school district, Andover, Massachusetts. AIRS is a classic, comprehensive individualized approach which uses programmed self-instructional materials. Almost all language arts instruction is delivered through the materials with teachers, aides, and parent volunteers giving one-to-one assistance as needed. Comparisons of Stanford Achievement Test scores in student cohorts before and after the AIRS program was introduced show clear gains which were maintained over several years. An AIRS evaluation in rural Rhode Island also found achievement gains each year after the program was installed. Again, no evidence is given to suggest that AIRS would be effective with disadvantaged or low-achieving students.

STAMM. Systematic Teaching and Measuring Mathematics (STAMM) was developed and initially evaluated in Jefferson County, Colorado, a Denver suburb. Like AIRS, STAMM is a fully individualized instructional program, which relies primarily on self-instructional materials and one-to-one instruction by teachers and aides. Evaluations in Jefferson County and in an Atlanta suburb showed that students who experienced STAMM scored higher on the CTBS than did students in earlier cohorts. Special forms of STAMM exist for use with Chapter I and learning disabled students, but evaluations with these or other at-risk students are not known to exist.

Other Individualized Programs. What is noteworthy about Table 2 is not so much the programs listed there as the programs not listed. A large number of JDRP-approved programs used individualized models, and the broader educational literature has many studies of such methods. Yet very few of these present convincing evidence of effectiveness.

One individualized program which was carefully considered for inclusion in Table 2 is the University of Kansas Behavior Analysis Program (Ramp & Rhine, 1981). This program is generally acknowledged to have been the second most effective of the Planned Variation Follow Through models, after Distar. Also, two Behavior Analysis sites, in Trenton, New Jersey (JDRP No. 77-139) and Waukegan, Illinois (JDRP No. 77-126) passed JDRP review with convincing, control-group evaluations (conducted as part of the overall Abt evaluation). However, site variability for Behavior Analysis was extreme; in some locations, especially New York City, Behavior Analysis students scored significantly worse than non-Follow Through students. Overall, Kennedy (1978) reports effect sizes of -.12 for reading comprehension and +.02 for language. Only in math computations is there positive evidence (ES=+.28), but math problem solving scores slightly favored the control group (ES=-.06).

Another individualized model, the Responsive Early Childhood Education Program (RECEP; JDRP No. 77-154) also presents a paradox in that data submitted to the JDRP show some positive effects but the overall combined results from all Planned Variation sites did not. Several other individualized programs compared experimental to "matched" control groups without establishing that the control groups were initially equivalent (e.g., Randolph County Follow Through, JDRP No. 77-149), or presented only minimal spring-to-spring gains (e.g, Personalized Classroom Management System, JDRP No. 78-170).

The JDRP programs which were excluded from Table 2 rarely provide data that could indicate that the programs are <u>not</u> effective, as such programs

would be unlikely to pass JDRP review. However, in the broader education literature there have been several good-quality evaluations of individualized programs which have indicated few positive effects (see Horak, 1971; Miller, 1976; Thompson, 1975). When positive effects of individualized programs have been found, they are often in studies with poor methodological characteristics. For example, Wang and Walberg (1983) claimed positive effects of the University of Pittsburgh's Adaptive Learning Environments Model (ALEM) presenting comparisons of "expected" scores based on national norms to scores of students in the ALEM program. However, the one study of this method to use a control group found no achievement differences (Wang & Birch, 1984).

Table 3 Here

Cooperative Learning

Table 3 lists two cooperative learning programs, TAI and CIRC, both of which were developed and evaluated at the Johns Hopkins University Center for Research on Elementary and Middle Schools. All four of the studies which evaluated these models (plus four additional studies which had durations of less than a semester) used control group designs, including the only two studies in Tables 1-3 in which teachers were randomly assigned to treatments (Slavin & Karweit, 1985).

TAI. Team Accelerated Instruction or TAI (Slavin, 1985) combines cooperative learning with a continuous progress approach to mathematics

instruction. In this model, students are assigned to 4-5 member, mixed-ability learning teams, and they are also assigned to skill-level groups on the basis of a placement test. Teachers instruct the skill level groups on the concepts of mathematics, after which students return to their teams and work on self-instructional materials. Teammates frequently check one another's work against answer sheets, assist one another with difficult problems, and prepare one another for quizzes. Quizzes are taken apart from the team area, and achieving mastery score on quizzes adds points to students' team scores. At the end of each week, teams which exceed a pre-established criterion in terms of numbers of units mastered may earn certificates or other rewards. Specific procedures and parallel assessments are provided for students who do not achieve at a mastery level.

TAI has been evaluated in six field experiments, but only three of these met the one-semester duration requirement for inclusion in the Tables. Two of the three studies (Slavin & Karweit, 1985) are the only studies listed in Tables 1-3 which randomly assigned teachers to treatments. One of these took place in and around Hagerstown, Maryland, and the other in inner-city Wilmington, Delaware. The Hagerstown study compared TAI to a control group and to the Missouri Mathematics program (Good, Grouws, & Ebmeier, 1983), a whole-class instructional method which incorporates teaching behaviors derived from process-product research. The Wilmington study compared TAI to the Missouri Mathematics program only; and a study in a Baltimore suburb (Slavin, Madden, & Leavey, 1984) compared TAI to an untreated control group.

Results of all three studies favored the TAI classes, with the largest effects in the Wilmington study. Separate analyses for mainstreamed

academically handicapped students in the suburban study also found positive effects for these students.

CIRC. Cooperative Integrated Reading and Composition (Madden et al., 1986; Stevens et al., in press) uses a combination of mixed-ability, cooperative work groups and skill-based reading groups to teach reading, language arts, and writing in the upper elementary grades. In CIRC, students are assigned to pairs within their reading groups, and then the pairs are assigned to 4-5 member teams, so that the teams are heterogeneous. During reading periods, teachers work with reading groups over the course of each week to set a purpose for reading, introduce vocabulary, discuss the characters, setting, problems, and problem solutions in narrative stories, and so on. Students back in their team areas work in pairs on a prescribed series of activities, including reading basal stories to one another, identifying elements of story structure, predicting story outcomes, practicing vocabulary, word lists, and spelling, and writing in response to stories. Each week teachers provide instruction in such reading comprehension skills as finding the main idea. Students earn points for their teams based on the sum of their individual performances on regular quizzes. The CIRC language arts program is based on a writing process model, with students working together in their teams to plan, draft, revise, edit, and ultimately publish compositions. Language mechanics instruction is provided in the context of students' writing.

The one semester-long study which evaluated CIRC found positive effects in comparison to matched control classes on CTBS reading comprehension, reading vocabulary, language expression, and language mechanics scales.

Separate analyses for mainstreamed academically handicapped students found strong effects on both reading scales but not on the language measures; for remedial reading students (mostly Chapter I), effects were similar to those in the full sample.

Other Research on Cooperative Learning. While the four studies cited above are the only ones to use standardized reading and math measures at the elementary level in studies of at least a semester, several additional studies support the achievement effects of cooperative learning. Two shorter studies of TAI found positive effects on math computations measures (Slavin, Leavey, & Madden, 1984), though a third found no differences (Oishi, Slavin, & Madden, 1983). A twelve-week study of CIRC (Madden, Stevens, & Slavin, 1986) found positive effects on CTBS reading comprehension, reading vocabulary, language expression, and spelling scales. Other cooperative learning methods, such as Student Teams-Achievement Divisions (STAD) and Teams-Games-Tournaments (TGT) have had positive effects in mathematics, language arts, and other subjects at the elementary and secondary levels (Slavin, 1983a, b). In addition, all of the cooperative learning methods have had positive effects on such outcomes as race relations, acceptance of mainstreamed students, and self-esteem (Slavin, 1983a).

Other Classroom Models

Considering that there are more than 400 programs passed by the JDRP, plus many others evaluated and reported in the educational literature, how is it that only sixteen programs in three categories are emphasized here?

Essentially, this paper took a conservative approach in identifying effective programs. Only programs with convincing evidence of effectiveness were listed, and even these are not all certain, as there is always the possibility that data submitted to JDRP or published in journals is taken from one or more selected sites and is not representative of all program sites. Yet it seems certain that there are more than sixteen effective classroom programs.

This section discusses the state of the evidence on well known or widely used models not represented among those listed in Tables 1-3.

Developmental/Humanistic Models. One major category of programs not listed in this paper is models that might be categorized as "developmental/humanistic." This includes the open classroom, Piagetian-based models, and other types of programs in which students are given choices of activities appropriate to their developmental levels, and are encouraged to discover and experience language and mathematical concepts rather than to master a pre-established series of skills.

One important example of the developmental/humanistic approach is the program developed by the Bank Street College of Education (Gilkeson, Smithberg, Bowman, & Rhine, 1981). Several sites using the Bank Street Program passed JDRP review (JDRP No. 77-156). However, the overall effects of this program in the Planned Variation evaluation were about zero in reading and language and negative in math. The High Scope/Cognitively Oriented Curriculum (Weikart, Hohmann, & Rhine, 1981), a child-centered, Piagetian model which extends into the elementary grades, many of the elements of the Perry Preschool Program (Weikart, 1974), also has several

JDRP-approved sites (JDRP No. 77-123), but came out near the bottom of the list in the overall Abt evaluations (Kennedy, 1978). What these findings may imply is that the developmental/humanistic models can be effective, as evidenced by their apparent success in several sites, but the evidence for this or evidence of what is required to ensure their success are essentially lacking. Also, it should be noted that critics of the Abt evaluation (e.g., House, Glass, McLean, & Walker, 1978) have argued that the use of standardized achievement tests may have biased the evaluation in favor of such highly structured programs as Distar and Behavior Analysis. However, the Abt evaluations focused on children who had been in their respective models three or four years. If they had any important effects on students' cognitive performance, a standardized test should have registered some (if not all) of those effects.

Group-Based Mastery Learning. Another widely used approach to school improvement is group-based mastery learning (Block, 1984; Block & Anderson, 1975; Guskey & Gates, 1985). The teacher assesses student learning at the end of a series of lessons. Those who achieve at a pre-established level of mastery (e.g., 80%) are given enrichment activities, while others are given corrective instruction designed to bring them up to the mastery criterion. Claims of substantial positive effects of group-based mastery learning on student achievement are based largely on very brief experiments. In a review of research on group-based mastery learning, Slavin (in press a) found that in studies of at least four weeks' duration there were no significant positive effects of mastery learning on standardized measures, and modest effects (median ES=+.25) on experimenter-made measures. However, it is important to note that the concepts of mastery assessment and

corrective instruction are central to all three categories of effective programs emphasized in this paper, and some of the continuous-progress programs, such as GEMS and the Outcomes-Driven Developmental model, identify themselves as mastery learning models (see Levine, 1985).

Principles of Effective Classroom Programs for Students at Risk

The purpose of this paper is not so much to identify individual programs as to discover principles which underlie effective models for students at risk. The effective programs identified in an extensive review of the literature fell into three categories: continuous progress, individualized instruction, and cooperative learning. However, the three effective individualized instruction models were all developed and evaluated in middle class, non-disadvantaged areas in which few students could be categorized as "at risk" for learning problems. Also, the broader literature has not tended to support the use of individualized, programmed instructional models. AIRS, STAMM, and the Matteson 4-D models may have unique features which could make them effective with at-risk students, but given the past history of research on individualized programs, this remains to be seen.

It is in the remaining two categories that the greatest confidence can be placed. In the case of the continuous progress models it is important to note not only the quality of the individual studies but also the number of successful models which began from very different theoretical bases and ended up with similar programs. Some of these models describe themselves as "individualized" (e.g., U-SAIL), but in order to provide adequate instruction to all students, incorporated groupings of students according to skill levels. Others (e.g., GEMS, ODDM) describe themselves as "mastery

learning," but incorporated subgrouping of students to accommodate instruction to individual needs. Developers of Distar describe the model as "direct instruction," yet the skill-level grouping, frequent assessment, and progress through a structured hierarchy of skills have little in common with more recent class-paced forms of "direct instruction" such as the Missouri Mathematics Program (Good et al., 1983).

Although very different in many respects, the effective cooperative learning programs listed in Table 3 have many features in common with the continuous progress models. First, instruction in TAI and CIRC is directed toward small, homogeneous groups of students, so instruction is at a level appropriate to student needs. Second, instruction comes from the teacher, not from written materials or peers. Third, students' prgress is constantly assessed and specific procedures exist to remediate any small problems before they become large ones. Finally, in TAI and in the reading comprehension component of CIRC, there is a structured hierarchy of skills which students must master. Cooperative learning methods add to these the opportunity for students to explain concepts to one another and help one another study, team incentives based on the learning of team members, and student responsibility for many elements of classroom management. This last feature may be particularly important in today's lean times; TAI and CIRC have always been evaluated in classrooms in which there are no aides or volunteers, while continuous progress and individualized models have almost always had additional personnel.

What emerges from an examination of the effective classroom programs is the following characterization: Consistently effective classroom programs

accommodate instruction to individual needs while maximizing direct instruction, and they frequently assess student progress through a structured hierarchy of skills.

The first part of this statement, the need to accommodate instruction to individual needs while maximizing direct instruction, is also supported by research on ability grouping in elementary schools, recently reviewed by Slavin (in press b). This review found that while assigning students to self-contained classes on the basis of general ability was ineffective, two forms of ability grouping did have considerable support in good-quality experimental studies. One was the Joplin Plan, in which students are regrouped for reading across grade lines into reading classes all operating at one reading level. This plan obviates the need for grouping within the reading class, giving teachers the opportunity to spend all of the reading period in direct instruction. Studies of the Joplin Plan and closely related forms of non-graded plans found consistent positive effects on standardized reading measures (median ES=+.44).

The second form of ability grouping found in the Slavin (in press b) review to be instructionally effective was within-class ability grouping in math (median ES=+.34). Use of math groups, especially if the number of groups is kept to two or three, strikes a balance between accommodating instruction to individual needs and providing adequate direct instruction. Since there is a need for independent seatwork in math, grouping within the class to meet students' diverse needs and having the teacher present two or three lessons at different levels may be the most efficient use of instructional time.

Slavin (1987) has argued that to make a meaningful difference in student achievement, four elements of classroom organization must be simultaneously addressed: Quality of instruction, appropriate level of instruction, incentive, and time. The continuous progress and cooperative learning methods address most of these elements, in particular in providing appropriate levels of instruction (by grouping students according to skill level) while maintaining an adequate quality of instruction (by having the teacher be responsible for direct instruction and by having a well thought out sequence of instructional objectives). Cooperative learning adds to the incentive element the use of team rewards based on team members' learning, and both types of programs emphasize effective use of time through careful structuring of classroom activities.

The importance of accommodating student needs while maintaining adequate direct instruction is perhaps greatest for at-risk students. Students who begin to experience difficulties in their early years in school are unlikely to be able to keep up with the class pace or grade level expectations. In response to this, schools provide remedial and special education for students who are falling behind. Yet these programs rarely accelerate students enough to enable them to catch up with their classmates (Madden & Slavin, 1983; Carter, 1984). Building classroom models able to meet diverse needs would seem to be a better approach to the problems of at-risk students than creating a parallel instructional system, which creates a large set of new problems in terms of coordination with the regular program, transitions, and so on (Johnston et al., 1985).

Conclusions

This paper reviewed all sources of evidence concerning classroom programs which have been or could be applied to accelerating the reading and math achievement of at-risk students. Among programs whose evaluations used control group designs with good evidence of initial experimental-control equivalence, study durations of at least a semester, and standardized measures of reading and math, two categories of programs emerged as particularly effective: Continuous progress and cooperative learning. These approaches have in common the use of instruction to small, skill-based groups; instruction from the teacher rather than from materials or peers; and frequent assessment of student progress through a hierarchical curriculum.

However, while this review is based on the best evidence currently available concerning the achievement effects of practical classroom instructional models, it is important to reiterate that this evidence is far from definitive. Even the carefully selected programs emphasized in this review could prove to be ineffective in later research; as noted previously, it is always possible that data submitted to the JDRP or to journals is from the <u>best</u> sites, not from a representative sample of implementations. It is also likely that many programs excluded from this review will prove to be effective or could be modified to be effective. Further, while it is of some use to know that Program X is effective in comparison to traditional methods, there is much more we must understand to make this information useful in informing us about program design and classroom instruction in general. What elements of Program X account for its effects? For whom is it effective? Does it have important side effects?

We are still at a primitive state in understanding effective instruction for students at risk. What is needed is a renewed focus on development, evaluation, and component analyses of programs based on many models of teaching and learning.

One approach to identifying effective, transportable instructional models for students at risk would be learning from the mistakes of the Planned Variation Follow Through evaluation but trying again to accomplish its objectives. A very small proportion of the enormous Chapter I budget, for example, could fund development of promising models, evaluation and component analyses of such models, and ultimately evaluation of the models in new sites with multiple measures (not only standardized tests), with random assignment of programs to schools. In this way, we could finally achieve the ambitious goals of Chapter I to make a meaningful difference in the life chances of disadvantaged students.

While we cannot conclude that the search for effective classroom programs for at-risk students is over, the findings of this review do have important implications for current practice. It identified several programs which have demonstrated potential for increasing the achievement of disadvantaged and low achieving students and identified program elements which seem to characterize effective models. The identified programs are all designed to be replicable and transportable; in most cases the developers have training staffs who work to help school districts successfully implement the models. However, there are two notes of caution that practitioners should bear in mind. First, all of the effective models identified in this review are complex, and are quite different from usual practice. Adequate provisions

ERIC

for training, followup, and monitoring of project implementations is imperative. Any of these programs can fail (and most have failed) when inadequate resources have been devoted to implementation. Program implementations should start on a small scale, beginning with volunteers and only gradually expanding from a solid base of success. Second, districts should conduct their own evaluations of new programs, using random assignment or careful matching to compare program users and non-users. The fact that a program was effective elsewhere suggests but does not guarantee that it will be successful in any particular district.

The search for reliably effective programs for students who are at risk for school failure is a task of great importance. This paper and companion papers by Madden and Slavin (1987) and Karweit (1987) are intended to sum up where we are now, in the hope that we can learn from the past in helping direct future research, development, and evaluation into effective instruction for our neediest students.

REFERENCES

Allington, R.L. (in press). Designing effective compensatory reading instruction: A review. Elementary School Journal.

Archambault, F.X. (1987, April). Pullout versus inclass instruction in compensatory instruction. Paper presented at the Annual Convention of the American Educational Association, Washington, DC.

Becker, W.C., & Carnine, D. (1980). Direct Instruction: An effective approach for educational intervention with the disadvantaged and low performers. In B.J. Lahey & A.E. Kazdin (eds.), Advances in child clinical psychology (pp. 429-473). New York: Plenum.

Becker, W.C., Engelmann, S., Carnine, D.W., & Rhine, W.R. (1981). Direct Instruction Model. In W.R. Rhine (ed.), Making schools more effective: New directions from follow through (pp. 95-154). New York: Academic Press.

Birman, B.F. (1981). Problems of Overlap Between Title I and P.L. 94-142: Implications for the federal role in education. Educational Evaluation and Policy Analysis, 3, 5-19.

Block, J.H., & Anderson, L.W. (1975). Mastery learning in classroom instruction. New York: Macmillan.

Bloom, B.S. (1984). The 2 sigma problem: The search for methods of instruction as effective as one-to-one tutoring. Educational Reserach, 13, 4-16.

Brophy, J.E., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (ed.), Handbook of research on teaching (Third Edition, pp. 328-375). New York: Macmillan.

Calfee, R. (1986, May). Compensatory Reading. Paper presented at the OERI Conference on Effects of Alternative Designs in Compensatory Education, Washington, DC.

Carter, L.F. (1984). The sustaining effects study of compensatory and elementary education. Educational Researcher, 13(7), 4-13.

Cooper, H.M. (1984). The integrative research review: A systematic approach. Beverly Hills, CA: Sage.

Coulson, J.E., Hanes, S.C., Ozenne, D.G., Bradford, C., Doherty, W.J., Duck, G.A., & Hemenway, J.A. The third year of emergency school aid act (ESAA) implementation. Santa Monica, CA: System Development Corporation.

Ellson, D.G. (1986). Improving the productivity of teaching: 125 exhibits. Bloomington, IN: Phi Delta Kappa.

Gabriel, R.M., Anderson, B.L., Benson, G., Gordon, S., Hill, R., Pfannenstiel, J., & Stonehill, R.M. (1985). Studying the sustained achievement of Chapter 1 students. Washington, DC.: U.S. Department of Education.

Gilkeson, E.C., Smithberg, L.M., Bowman, G.W., & Rhine, W.R. (1981). Bank Street Model: A developmental-interaction approach. In W.R. Rhine (ed.), Making schools more effective: New directions from follow through (pp. 249-190). New York: Academic Press.

Glass, G.V., & Smith, M.L. (1977). Pull out in compensatory education. Washington, DC.: Department of Health, Education, and Welfare.

Glass, G., McGaw, B., & Smith, M.L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Good, T., Grouws, D., & Ebmeier, H. (1983). Active mathematics teaching. New York: Longman.

Goodlad, J.I., & Anderson, R.H. (1963). The nongraded elementary school (rev. ed.). New York: Harcourt, Brace, & World.

Griswold, P.A., Cotton, K.J., & Hansen, J.B. (1986). Effective compensatory education sourcebook. Washington, DC: U.S. Government Printing Office.

Guskey, T.R., & Gates, S.L. (1985, April). A synthesis of research on group-based mastery learning programs. Paper presented at the Annual Convention of the American Educational Research Association, Chicago.

Hales, C. (1983a). A teacher's guide to individualizing reading. Salt Lake City, UT: U-SAIL Project.

Hales, C. (1985b). Teaching for mastery in mathematics. Salt Lake City, UT: U-SAIL Project.

Horak, V.M. (1981). A meta-analysis of research findings on individualized instruction in mathematics. Journal of Educational Research, 74, 249-253.

House, E., Glass, G., McLean, L.D., & Walker, D. (1978). No simple answer: Critique of the follow through evaluation. Harvard Educational Review, 48, 128-160.

Johnston, P., Allington, R., & Afflerbach, P. (1985). Curriculum confluence in classroom and clinic. Elementary School Journal, 85, 465-477.

Kaestle, C.F., & Smith, M.S. (1982). The historical context of the federal role in education. Harvard Educational Review, 52, 383-408.

Karweit, N.L. (1987, April). Effective early childhood education and kindergarten programs for children at risk of academic failure. Paper presented at the Annual Convention of the American Educational Research Association, Washington, DC.

Kelly, F.J., Veldman, D.J., & McGuire, C. (1964). Multiple discriminant prediction of delinquency and school dropouts. Educational and Psychological Measurement, 24, 535-544.

Kennedy, M. (1978). Findings from the Follow Through planned variation study. Educational Researcher, 7, 3-11.

ERIC

Kennedy, M.M., & Birman, B.F. (1986). The effectiveness of Chapter 1 services. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Kimbrough, J., & Hill, P.T. (1981). The aggregate effects of federal education programs. Santa Monica, CA: Rand Corp.

Lee, G.V., & Rowan, B. (1986). The management and delivery of instructional services to Chapter 1 students: Case studies of twelve schools. San Francisco: Far West Laboratory for Educational Reserach and Development.

Levine, D.U. (ed.) (1985). Improving student achievement through mastery learning programs. San Francisco: Jossey-Bass.

Light, R.J., & Pillemer, D.B. (1984). Summing Up: The science of reviewing research. Cambridge, MA: Harvard University Press.

Madden, N.A., & Slavin, R.E. (1983). Mainstreaming students with mild academic handicaps: Academic and social outcomes. Review of Educational Research, 53, 519-569.

Madden, N.A., & Slavin, R.E. (1987, April). Effective pull-out programs for students at risk. Paper presented at the Annual Convention of the American Educational Research Association, Washington, DC.

Madden, N.A., Slavin, R.E., & Stevens, R.J. (1986). Cooperative Integrated Reading and Composition: Teacher's manual. Baltimore, MD.: Johns Hopkins University Center for Research on Elementary and Middle Schools.

Meyer, L.A. (1984). Long-term academic effects of the direct instruction project Follow Through. Elementary School Journal, 84, 380-394.

Miller, R.L. (1976). Individualized instruction in mathematics: A review of research. The Mathematics Teacher, 69, 345-351.

Novak, C.D. (1975, April). An unobstrusive experimental evaluation of a systems approach for teaching reading. Paper presented at the annual convention of the American Educational Research Association, Washington, DC.

Ramp, E.A., & Rhine, W.R. (1981). Behavior analysis model. In W.R. Rhine (ed.), Making schools more effective: New directions from follow through (pp. 155-200). New York: Academic Press.

Rowan, B., Guthrie, L.F., Lee, G.V., & Guthrie, G.P. (1986). The design and implementation of Chapter 1 services: A study of 24 schools. San Francisco: Far West Laboratory for Educational Research and Development.

Sargent, L.R. (1981). Resource teacher time utilization: An observational study. Exceptional Children, 47, 420-425.

Savage, D.G. (1987). Why Chapter 1 hasn't made much difference. Phi Delta Kappan, 68, (8), 581-584.

Slavin, R.E. (1983a). Cooperative learning. New York: Longman.

Slavin, R.E. (1983b). When does cooperative learning increase student achievement? Psychological Bulletin, 94, 429-445.

Slavin, R.E. (1985). Team-assisted individualization: combining cooperative learning and individualized instruction in mathematics. In R.E. Slavin, S. Sharan, S. Kagan, R. Hertz-Lazarowitz, C. Webb, & R. Schmuck (eds.), Learning to cooperate, cooperating to learn (pp. 177-209). New York: Plenum.

Slavin, R.E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. Educational Research, 15(9), 5-11.

Slavin, R.E. (1987). A theory of school and classroom organization. Educational Psychologist, 22, 89-108.

Slavin, R.E. (in press a). Mastery learning reconsidered. Review of Educational Research.

Slavin, R.E. (in press b) Ability grouping and student achievement in elementary schools: A best-evidence systhesis. Review of Educational Research.

Slavin, R.E., & Karweit, N.L. (1985). Effects of whole-class ability grouped, and individualized instruction on mathematics achievement American Educational Research Journal, 22, 351-367.

Slavin, R.E., Leavy, M., & Madden, N.A. (1984). Combining cooperative learning and individualized instruction: Effects on student mathematics achievement, attitudes, and behaviors. Elementary School Journal, 84, 409-422.

Slavin, R.E., Leavey, M.B., & Madden, N.A. (1986). Team Accelerated Instruction - Mathematics. Watertown, MA: Mastery Education Corporation.

Slavin, R.E., Madden, N.A., & Leavey, M. (1984). Effects of team assisted individualization on the mathematics achievement of academically handicapped and nonhandicapped students. Journal of Educational Psychology, 76, 813-819.

Stebbins, L.B., St. Pierre, R.G., Proper, E.C., Anderson, R.B., & Cerva, T.R. (1977). Education as experimentation: A planned variation model (vols. IV A-D). Cambridge, MA: Abt Associates.

Stevens, R.J., Madden, N.A., Slavin, R.E., & Farnish, A.M. (in press). Cooperative Integrated Reading and Composition: Two field experiments. Reading Research Quarterly.

Stroup, A.L., & Robins, L.N. (1972). Elementary-school predictors of school dropout among black males. Sociology of Education, 45, 212-222.

Thompson, R. (1975). Individualized reading: A summary of research. Educational Leadership, 32, 57-63.

Vanecko, J., Ames, N., & Archambault, F. (1980). Who benefits from federal education dollars? Cambridge, MA: ABT Books.

Wang, M.C., & Birch, J.W. (1984). Comparison of a full-time mainstreaming program and a resource room approach. Exceptional Children, 51, 33-40.

Wang, M.C., & Walberg, H.J. (1983). Adaptive instruction and classroom time. American Educational Research Journal, 20, 601-626.

Weikart, D.P. (1974). Curriculum for early childhood education. Focus on Exceptional Children, 6, 1-8.

Weikart, D.P., Hohmann, C.F., & Rhine, W.R. (1981). High/scope cognitively oriented curriculum model. In W.R. Rhine (ed.), Making schools more effective: New directions from follow through (pp. 201-248). New York: Academic Press.

Table 1
Continuous Progress Programs

Program and Source	Grades Subjects	Description	Evaluation	Effect Sizes	
I. Programs Evaluated Using Control Group Designs					
DISTAR (Becker & Carnine, 1980; Abt; JDRP #77-122, JDRP #80-50)	K-6 Reading Math	Teacher instructs small groups using highly structured, scripted lessons. Students frequently assessed, regrouped.	Abt evaluation found positive effects in rdg and math in most sites; JDRP-approved sites include Flint, MI; Dayton, OH; Flippin, AK; E. Las Vegas, NM; Uvalde, TX; Kingstree, SC; Cherokee, NC; Washington, DC.	Gr. K-3 (4 yrs.)	Rdg. Comp. +.07 Language +.84 Math Comp. +.57 Math P.S. +.17
U-SAIL (JDRP #76-95)	1-9 Reading Math	Combines continuous progress, individualized activities. Adapts to existing curriculum, materials.	Both time series and control group comparisons show clear effects in reading and math in suburban school near Salt Lake City. <u>Not disadvantaged</u>.	Gr. 3-5 (2 yrs.)	Reading +.4[illegible] Math +.27
PEGASUS-PACE (JDRP #1, JDRP #79-1)	K-8 Reading	Students proceed through 17 rdg levels, gr. K-8; frequently assessed & regrouped.	Matched control group design found convincing effects in reading in rural Princeton, IL; some students were Title I, but mostly <u>not disadvantaged</u>.	Gr. 2 +.72 Gr. 3 +.80 Gr. 4 +.50 Gr. 5 +.39 Gr. 6 +.43	Rdg. Voc. +.53 Rdg. Comp. +.61
ECRI (JDRP #74-48)	1-6 Reading	Teacher instructs small groups; frequent mastery checks	Most evaluations fall-spring or spring-spring. One control group evaluation in suburban Chapter I school near Cincinnati found convincing effects.	Gr. 4	Rdg. Voc. +.51 Rdg. Comp. +.39
Project INSTRUCT (JDRP #75-37)	K-3 Reading	Cross-grade grouping according to rdg skills; students proceed through skills at own rate.	Compared students in Lincoln, NE, schools that had successfully implemented program to matched schools. <u>Not disadvantaged</u>.	Gr. 2	Word. Knwl. +.30 Reading +.23
GEMS (JDRP #79-2)	K-12 Reading	Students work in small groups or individually on materials at their own level. Frequent assessment, mastery tests, corrective instruction.	Compared project schools with matched controls in suburb of Salt Lake City <u>Not disadvantaged</u>.	Gr. 1 +.11 Gr. 2 +.00 Gr. 3 +.41 Gr. 4 +.23 Gr. 5 +.44 Gr. 6 +.39	Rdg. Voc. +.[illegible]9 Rdg. Comp. +.14
Early Childhood Preventative Curriculum (JDRP #74-57)	1 only Reading	Students identified as high risk given intensive continuous progress, diagnostic-prescriptive program with small group and individualized activities.	Compared project schools to matched controls in Miami, FL.	Gr. 1	Par. Mng. +.95 WRAT +.28

ERIC

Program and Source	Grades Subjects	Description	Evaluation
II. Programs Evaluated Using Year-to-Year Gains			
WILASD (JDRP #83-2)	1-3 Reading	Combines continuous progress, whole class, and individualized instruction.	Compared cohorts in mostly Hispanic, low income school in Weslaco, TX before program begun and and then several years after. Clear differences in scores.
COMP (JDRP #74-114)	1-8 Math	Students taught ir small groups, progress tnrough 25 levels in 8 yrs.	Fall-to-fall trends in NCE's show clear gains in math. Columbia, MO - Not disadvantaged.
CLIMB (JDRP #81-44)	1-12 Reading Math	Diagnostic-prescriptive, continuous progress. Title I tchrs mostly consult, do some pullout.	Grade 4-6 Title I students in suburban Middlesex, NJ gained significantly in fall-to-fall rdg and math NCE's.
Outcomes-Driven Developmental Model (Mamary & Rowe, 1985; JDRP #85-7)	1-8 Reading Math	Students grpd across grades according to skills. Frequent mastery assessments, corrective instruction. Uses cooperative learning during independent practice.	Gains over several years in percent of students scoring above grade level, Johnson City, NY. Mostly not disadvantaged.

ERIC

Table 2
Individualized Instruction Programs

Program and Source	Grades Subjects	Description	Evaluation	Effect Sizes	
I. Programs Evaluated Using Control Group Designs					
Matteson Four-Dimensional Reading Program (JDRP #77-109)	3-8 Reading	Students work on individualized learning packets 40-60% of rdg time.	Compared project classes to matched control classes in suburban Chicago. Not disadvantaged.	Gr. 3-6 (3 yrs.)	Reading +.57
II. Programs Evaluated Using Year-to-Year Gains					
AIRS (JDRP #74-25)	1-6 Reading	Students mostly work on individualized materials, some small group work.	Comparison of spring testing shows cohort gains in suburban Andover, MA. Not disadvantaged.		
STAMM (JDRP #76-87)	K-8 Math	Structured, individualized program with programmed materials.	Evaluations in suburban Denver and suburban Atlanta show gains in spring NCE's. Not disadvantaged.		

Table 3
Cooperative Learning Programs

Program and Source	Grades Subjects	Description	Evaluation	Effect Sizes		
I. Programs Evaluated Using Control Group Designs						
Team Accelerated Instruction (Slavin et al., 1984; Slavin & Karweit, 1985; JDRP #84-5)	3-6 Math	Students work on programmed materials in mixed-ability teams while tchrs teach same-ability teaching groups.	In three studies, two with random assignment, one matched, TAI classes exceeded control. One study in rural MD, one in suburban MD, one in urban DE.	Full Sample (Gr. 3-6)	Math Comp.	+.50
					Math C&A	+.06
				Mainstreamed Students	Math Comp.	+.38
					Math C&A	+.47
Cooperative Integrated Reading & Composition (Stevens et al., in press)	3-5 Reading Writing	Students work in mixed-ability teams while tchr teaches reading groups.	CIRC classes exceeded matched control classes in study in suburbs of Baltimore.	Full Sample (Gr. 3-4)	Rdg. Voc.	+.12
					Rdg. Comp.	+.35
					Oral Rdg.	+.54
					Language	+.30
				Remedial Students	Rdg. Voc.	+.26
					Rdg. Comp.	+.40
					Oral Rdg.	+.71
					Language	+.37
				Mainstreamed Students	Rdg. Voc.	+.90
					Rdg. Comp.	+.99
					Language	.00